UNIT-I

Computer Organization:

It refers to the operational units and their interconnections structures..

Computer hardware:

Consists of electronic circuits, displays, magnetic and optical storage media, electromechanical equipment and communication facilities.

Computer Architecture:

It is concerned with the design of instruction set, instruction format, addressing modes of the computer. It includes the information formats, the instruction set and techniques for addressing memory.

Functional Units

A computer consists of 5 main parts.

Input

Memory

Arithmetic and logic

Output

Control Units

Functional units of a Computer

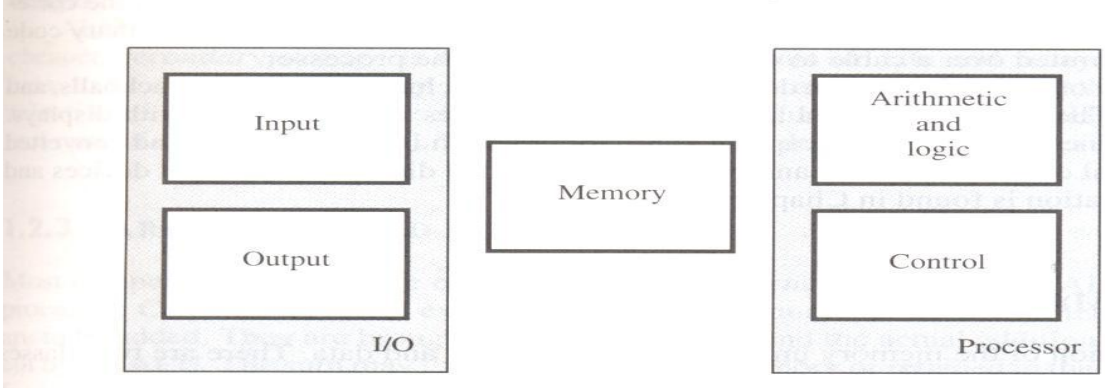


Figure 1.1 Basic functional units of a computer.

Input unit accepts coded information from human operators, from electromechanical devices such as keyboards, or from other computers over digital communication lines. The information received is either stored in the computer's memory for later reference or immediately used by the arithmetic and logic circuitry to perform the desired operations. The processing steps are determined by a program stored in the memory. Finally the results are sent back to the outside world through the output unit.

All of these actions are coordinated by the control unit. The list of instructions that performs a task is called a program. Usually the program is stored in the memory. The processor then fetches the instruction that make up the program from the memory one after another and performs the desire operations.

1.1 Input Unit:

- Computers accept coded information through input units, which read the data.
- Whenever a key is pressed, the corresponding letter or digit is automatically translated into its corresponding binary code and transmitted over a cable to either the memory or the processor.

Some input devices are

- ✓ Joysticks
- ✓ Trackballs
- ✓ Mouses
- ✓ Microphones (Capture audio input and it is sampled & it is converted into digital codes for storage and processing).

1.2.Memory Unit:

It stores the programs and data.

There are 2 types of storage classes

- ✓ Primary
- ✓ Secondary

Primary Storage:

It is a fast memory that operates at electronic speeds. Programs must be stored in the memory while they are being executed. The memory contains large no of semiconductor storage cells. Each cell carries 1 bit of information. The Cells are processed in a group of fixed size called Words. To provide easy access to any word in a memory, distinct address is associated with each word location. Addresses are numbers that identify successive locations. The number of bits in each word is called the word length. The word length ranges from 16 to 64 bits. There are 3 types of memory. They are

- ✓ RAM(Random Access Memory)
- ✓ Cache memory
- ✓ Main Memory

RAM:

Memory in which any location can be reached in short and fixed amount of time after specifying its address is called RAM. Time required to access 1 word is called Memory Access Time.

Cache Memory:

The small,fast,RAM units are called Cache. They are tightly coupled with processor to achieve high performance.

Main Memory:

The largest and the slowest unit is called the main memory.

1.3. ALU:

Most computer operations are executed in ALU.

Consider a example, Suppose 2 numbers located in memory are to be added. They are brought into the processor and the actual addition is carried out by the ALU. The sum may then be stored in the memory or retained in the processor for immediate use. Access time to registers is faster than access time to the fastest cache unit in memory.

1.4. Output Unit:

Its function is to send the processed results to the outside world. eg.Printer

Printers are capable of printing 10000 lines per minute but its speed is comparatively slower than the processor.

1.5. Control Unit:

The operations of Input unit, output unit, ALU are co-ordinate by the control unit.
The control unit is the Nerve centre that sends control signals to other units and senses their states.
Data transfers between the processor and the memory are also controlled by the control unit through timing signals.
The operation of computers are,

1. The computer accepts information in the form of programs and data through an input unit and stores it in the memory.
2. Information stored in the memory is fetched, under program control into an arithmetic and logic unit, where it is processed.
3. Processed information leaves the computer through an output unit.
4. All activities inside the machine are directed by the control unit.

BASIC OPERATIONAL CONCEPTS:

The data/operands are stored in memory.
The individual instruction are brought from the memory to the processor, which executes the specified operation.

EXAMPLE:1

Add LOC A ,R1

Instructions are fetched from memory and the operand at LOC A is fetched. It is then added to the contents of R0, the resulting sum is stored in Register R0.

EXAMPLE:2

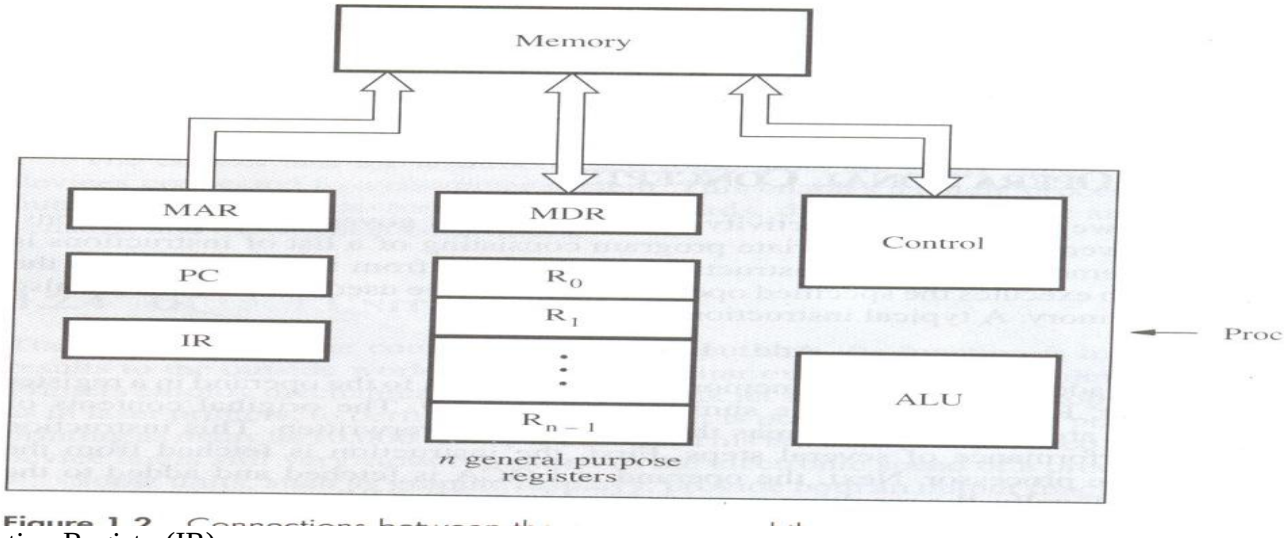
Load LOC A, R1
Transfer the contents of memory location A to the register R1.

EXAMPLE:3

Add R1 ,R0

Add the contents of Register R1 & R0 and places the sum into R0.

Fig:Connection between Processor and Main Memory



Instruction Register(IR)
Program Counter(PC)
Memory Address Register(MAR)
Memory Data Register(MDR)

Instruction Register (IR):

- It holds the instruction that is currently being executed.
- It generates the timing signals.

Program Counter (PC):

It contains the memory address of the next instruction to be fetched for execution.

Memory Address Register (MAR):

It holds the address of the location to be accessed.

Memory Data Register (MDR):

- It contains the data to written into or read out of the address location.
- MAR and MDR facilitates the communication with memory

Operation Steps:

- The program resides in memory. The execution starts when PC is point to the first instruction of the program.
- MAR read the control signal.
- The Memory loads the address word into MDR. The contents are transferred to Instruction register. The instruction is ready to be decoded & executed.

BUS STRUCTURES

A group of lines that serves as the connection path to several devices is called a Bus. A Bus may be lines or wires or one bit per line. The lines carry data or address or control signal.

There are 2 types of Bus structures. They are

1. Single Bus Structure
2. Multiple Bus Structure

3.1.Single Bus Structure:

- It allows only one transfer at a time.
- It costs low.
- It is flexible for attaching peripheral devices.
- Its Performance is low.

3.2.Multiple Bus Structure:

- It allows two or more transfer at a time.
- It costs high.
- It provides concurrency in operation.
- Its Performance is high.

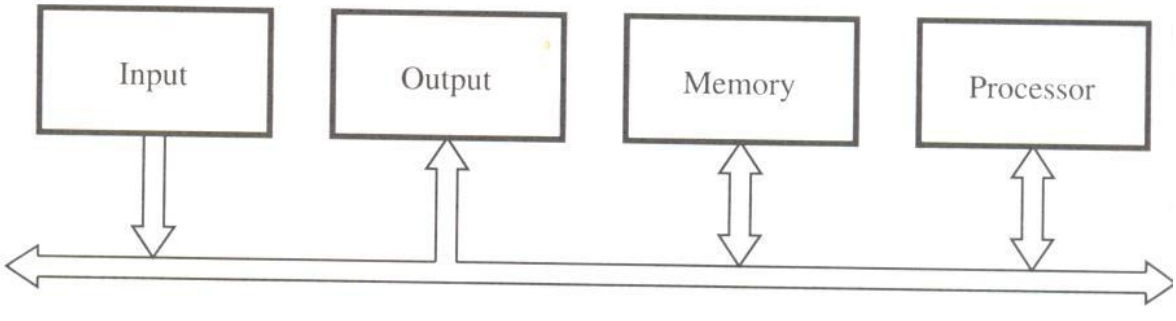


Figure 1.3 Single-bus structure.

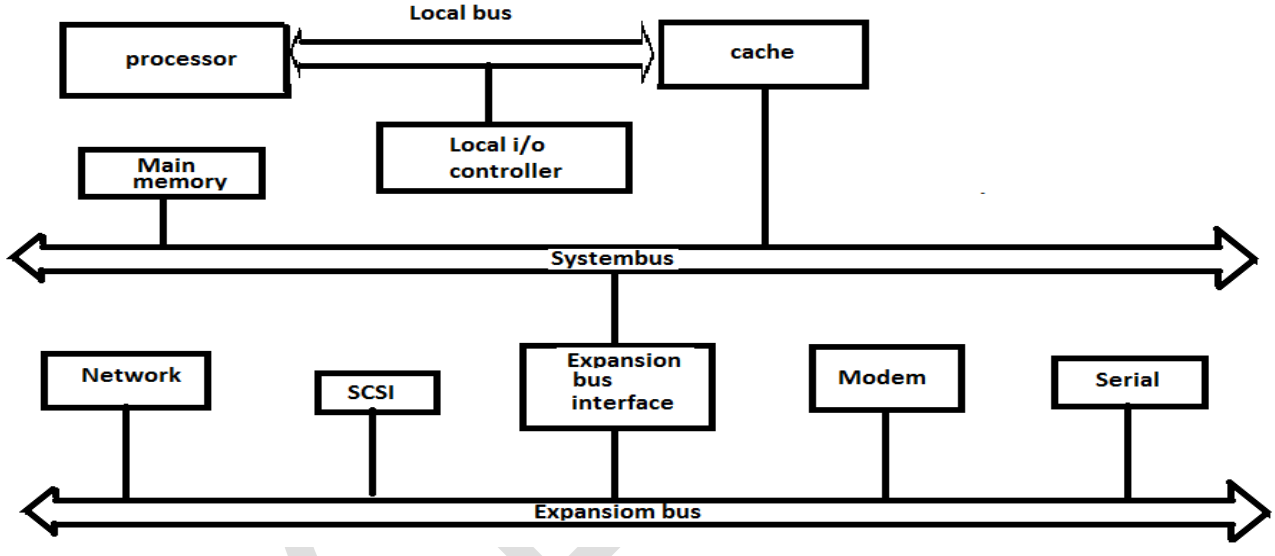
Devices Connected with Bus	**Speed**
Electro-mechanical decvices (Keyboard,printer)	Slow
Magnetic / optical disk	High
Memory & processing units	Very High

The Buffer Register when connected with the bus, carries the information during transfer.
The Buffer Register prevents the high speed processor from being locked to a slow I/O device during a sequence of data transfer.

Multiple bus Hierarchies:

There are two different types

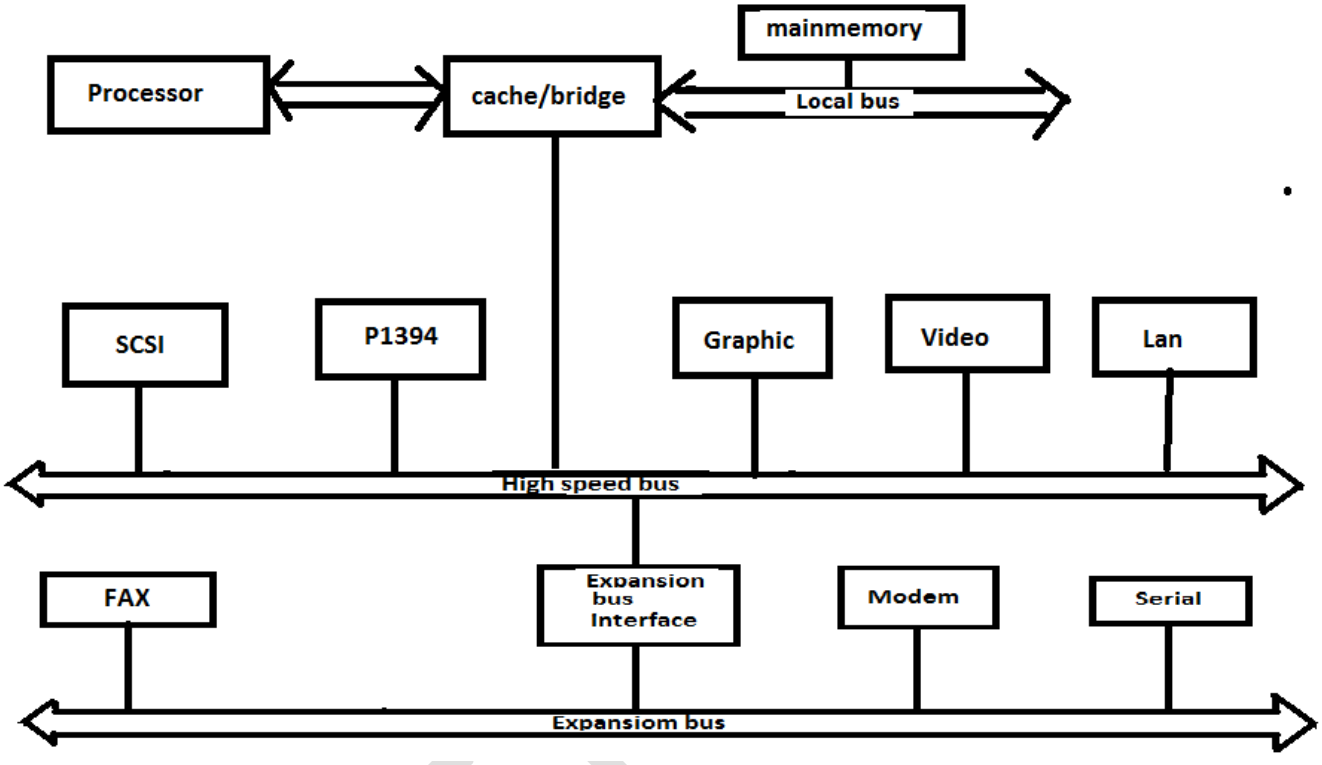
1. **Traditional –Bus architectures**
2. **High-performance architecture**



There is a local bus that connects the processor to a cache memory and that may support one or more local i/o devices. The cache memory controller connects to the system bus also, which is attached to the main memory modules. It is possible to connect I/O controller directly into the system bus. The most efficient solution is to make use of one or more expansion buses.
An expansion bus interface buffers data between the system bus and I/O controller on the expansion bus. This arrangement allows the system to support wide variety of I/O devices and at the same time insulate memory to processor traffic from I/O traffic.
Some of the devices connected to the expansion bus are

1. Network connections includes LAN and WAN
2. Small computer system interface(SCSI) it self is a type of bus used to support local disk drives and other peripherals.
3. A serial port used to support printer or scanner.

This architecture is efficient but, begins to break down as higher performance is seen in I/O devices. In addition to the growing demands, a common approach taken by the industry is to build a high speed bus, requiring a bridge between a processor bus and high speed bus. This arrangement is known as Mezzanine architecture which is shown below.



SOFTWARE:

System Software is a collection of programs that are executed as needed to perform function such as,

- Receiving & Interpreting user commands.
- Entering & editing application program and storing them as files in secondary Storage devices.
- Managing the storage and retrieval of files in Secondary Storage devices.
- Running the standard application such as word processor, games, and spreadsheets with data supplied by the user.
- Controlling I/O units to receive input information and produce output results.
- Translating programs from source form prepared by the user into object form.
- Linking and running user-written application programs with existing standard library routines.

Software is of 2 types. They are

1. Application program
2. System program

Application Program:

It is written in high level programming language(C,C++,Java,Fortran) The programmer using high level language need not know the details of machine program instruction.

System Program:(Compiler, Text Editor, File)

Compiler:

It translates the high level language program into the machine language program.

Text Editor:

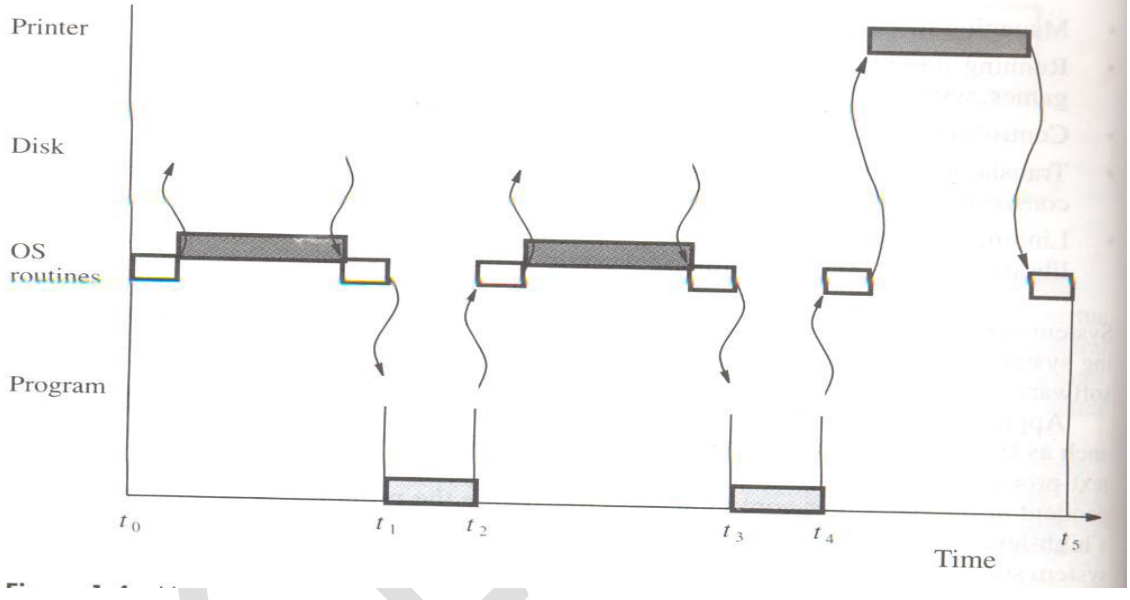
It is used for entering & editing the application program.

System software Component ->OS(OPERATING SYSTEM)

Operating System :
It is a large program or a collection of routines that is used to control the sharing of resource and interaction among various computer units.
Functions of OS:

- Assign resources to individual application program.
- Assign memory and magnetic disk space to program and data files.
- move the data between the memory and disk units.
- Handles I/O operation.

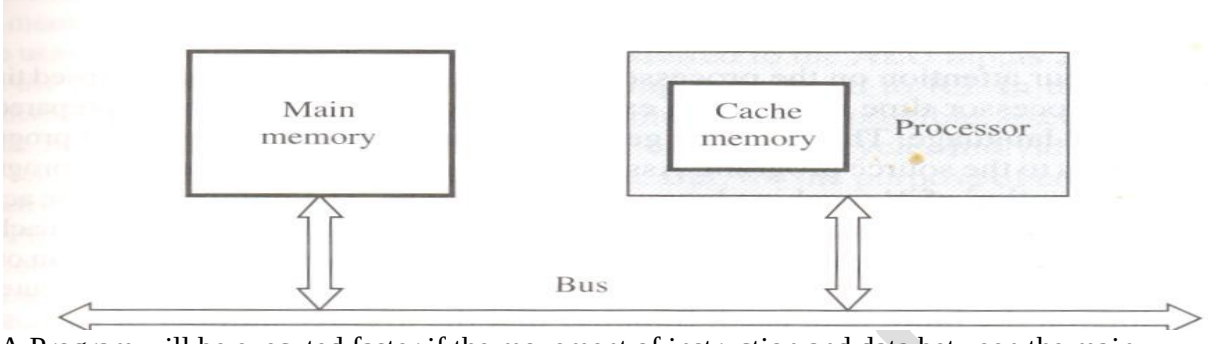
Fig:User Program and OS routine sharing of the process
Steps:



Steps:
1. The first step is to transfer the file into memory.
2. When the transfer is completed, the execution of the program starts.
3. During time period „t0" to „t1" , an OS routine initiates loading the application program from disk to memory, wait until the transfer is complete and then passes the execution control to the application program & print the results.
4. Similar action takes place during „t2" to „t3" and „t4" to „t5".
5. At „t5", Operating System may load and execute another application program.
6. Similarly during „t0" to „t1" , the Operating System can arrange to print the previous program"s results while the current program is being executed.
7. The pattern of managing the concurrent execution of the several application programs to make the best possible use of computer resources is called the multi-programming or multi-tasking

PERFORMANCE:
For best performance, it is necessary to design the compiler, machine instruction set and hardware in a co-ordinate way.
Elapsed Time→the total time required to execute the program is called the elapsed time.
It depends on all the units in computer system.
Processor Time→The period in which the processor is active is called the processor time.
It depends on hardware involved in the execution of the instruction.

Fig: The Processor Cache



A Program will be executed faster if the movement of instruction and data between the main memory and the processor is minimized, which is achieved by using the Cache.

Processor clock:

Clock→The Processor circuits are controlled by a timing signal called a clock.

Clock Cycle→The cycle defines a regular time interval called clock cycle.

Clock Rate, **R =1/P**

Where, P→Length of one clock cycle.

Basic Performance Equation:

T = (N*S)/R

Where, T→Performance Parameter

R→Clock Rate in cycles/sec

N→Actual number of instruction execution

S→Average number of basic steps needed to execute one machine instruction.

To achieve high performance,

N,S<R

Pipelining and Superscalar operation:

Pipelining→A Substantial improvement in performance can be achieved by overlapping the execution of successive instruction using a technique called pipelining.

Superscalar Execution →It is possible to start the execution of several instruction in everey clock cycles (ie)several instruction can be executed in parallel by creating parallel paths.This mode of operation is called the Superscalar execution.

Clock Rate:

There are 2 possibilities to increase the clock rate(R).They are,

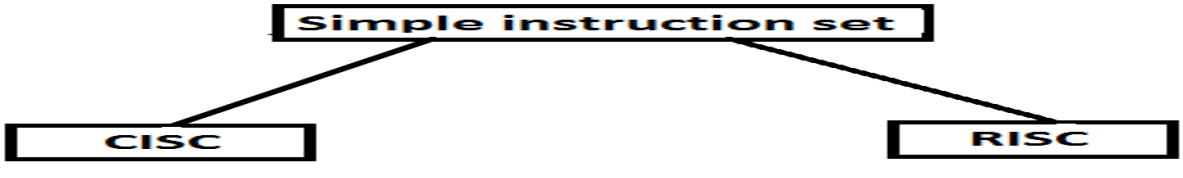
Improving the integrated Chip(IC) technology makes logical circuits faster.

Reduce the amount of processing done in one basic step also helps to reduce the clock period P.

Instruction Set:CISC AND RISC:

The Complex instruction combined with pipelining would achieve the best performance.

It is much easier to implement the efficient pipelining in processor with simple instruction set.



.

(Reduced Instruction Set Computer) **(Complex Instruction Set Computer)**

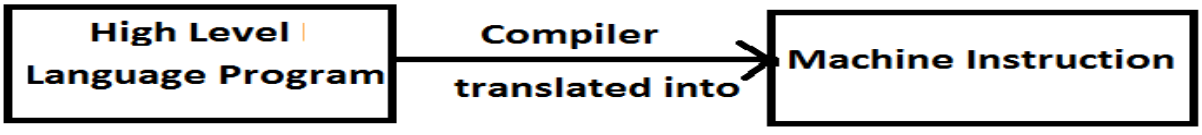
It is the design of the instruction set of a processor with simple instruction instruction

It is the design of the instruction set of a processor with complex

Functions of Compiler:

The compiler re-arranges the program instruction to achieve better performance.

The high quality compiler must be closely linked to the processor architecture to reduce the total number of clock cycles.



Performance Measurement:

The Performance Measure is the time it takes a computer to execute a given benchmark.

A non-profit organization called SPEC(System Performance Evaluation Corporation) selects and publishes representative application program.

$$\text{SPEC Rating} = \frac{\text{Running time on reference computer}}{\text{Running time on comuter under test}}$$

The Overall SPEC rating for the computer is given by,

SPEC Rating $= \left(\prod_{i=1}^{n} \text{SPEC}_i \right)^{1/n}$

EVOLUTION OF COMPUTERS

FIRST GENERATION (1945-1950)

1. Program and data reside in the same memory (stored program concepts – John von Neumann)
2. ALP was made used to write programs
3. Vacuum tubes were used to implement the functions (ALU & CU design)
4. Magnetic core and magnetic tape storage devices are used
5. Using electronic vacuum tubes, as the switching components

SECOND GENERATION (1955 – 1965)

1. Transistors were used to design ALU & CU
2. HLL is used (FORTRAN)

3. To convert HLL to MLL compiler were used
4. Separate I/O processor were developed to operate in parallel with CPU, thus improving the performance
5. Invention of the transistor which was faster, smaller and required considerably less power to operate

THIRD GENERATION (1965-1975)

1. IC technology improved
2. Improved IC technology helped in designing low cost, high speed processor and memory modules
3. Multiprogramming, pipelining concepts were incorporated
4. DOS allowed efficient and coordinate operation of computer system with multiple users
5. Cache and virtual memory concepts were developed
6. More than one circuit on a single silicon chip became available

FOURTH GENERATION (1975-1985)

1. CPU – Termed as microprocessor
2. INTEL, MOTOROLA, TEXAS,NATIONAL semiconductors started developing microproc essor
3. Workstations, microprocessor (PC) & Notebook computers were developed
4. Interconnection of different computer for better communication LAN,MAN,WAN
5. Computational speed increased by 1000 times
6. Specialized processors like Digital Signal Processor were also developed

BEYOND THE FOURTH GENERATION

(1985 – TILL DATE)

1. E-Commerce, E-banking, home office
2. ARM, AMD, INTEL, MOTOROLA
3. High speed processor -GHz speed
4. Because of submicron IC technology lot of added features in small size